

Wichtige statistische Begriffe

Der Stichprobenumfang n :

Oft ist es kaum möglich den Mittelwert für eine bestimmte Größe exakt zu berechnen, weil man den Wert nicht bei allen Untersuchungsobjekten bestimmen kann.

Bsp.: Um die mittlere Körpergröße aller Erwachsenen in Deutschland zu bestimmen, müsste man ALLE Erwachsenen Deutschland messen → quasi unmöglich!

Daher misst man die Körpergröße nur an einer ausgewählten Gruppe von Objekten, der **Stichprobe**. Die Anzahl der Objekte in dieser Gruppe heißt **Stichprobenumfang n** . Der Mittelwert der sich aus dieser Stichprobe errechnet, wird als Mittelwert aller Objekte (im obigen Beispiel: Alle Erwachsenen Deutschlands) angesehen.

Wichtig: Die Stichprobe muss **repräsentativ** sein, d. h.: Die Objekte der Gruppe müssen ZUFÄLLIG gewählt werden!

Bsp.: Bereits um 18.01 Uhr werden an einem Wahltag Hochrechnungen präsentiert, die nahezu immer sehr nahe an das spätere Ergebnis herankommen. Dies liegt daran, dass die Wahlforscher eine repräsentative Stichprobe an Wählern auswählen und diese befragen, welche Partei sie gewählt haben. Wichtig dabei ist, dass die Forscher zum Beispiel nicht **die ersten 10 Wähler aller Wahllokale** befragen. Es könnte ja z.B. sein, dass Frühaufsteher eher die Partei X wählen als diejenigen, die später aufstehen. Auch dürfen sie nicht nur Wähler der Wahllokale **in Berlin** befragen, weil diese vielleicht bevorzugt eine Partei Y wählen, die Landbewohner nicht wählen.

Auch darf die Stichprobe nicht zu klein sein! Wenn man nur 10 Leute befragt, welche Partei sie gewählt haben und nur einer zufällig angibt, die NPD gewählt zu haben, entsprächen das schon 10% der Stimmen. Umfragen oder wissenschaftliche Studien mit einer kleinen Stichprobe – unter Forschern sagt man: „Studien mit kleinem n “ – sind grundsätzlich sehr kritisch zu betrachten (Bsp.: Actimel stärkt das Abwehrsystem, $n=33$!)

Der Mittelwert \bar{x} (arithmetisches Mittel), \tilde{x} (Median):

Das **arithmetische Mittel** berechnet sich aus der Summe aller Einzelwerte geteilt durch die Anzahl der Objekte.

Bsp.: Ein Schüler erhält in Biologie im Laufe eines Schuljahres die Noten 3, 4, 2 und 3. Das arithmetische Mittel beträgt dann:

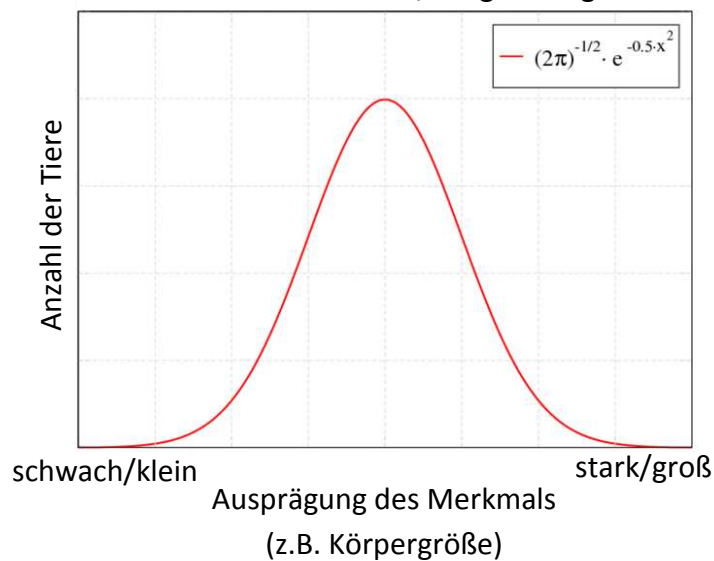
$$\frac{3 + 4 + 2 + 3}{4} = \frac{12}{4} = 3,00 \quad (\text{In der Statistik wäre das Berechnen des arithmetischen Mittels bei Schulnoten unzulässig!})$$

Der **Median** (auch: Zentralwert) gibt den Wert an, der eine beobachtete Zahlenreihe so teilt, dass über ihm genau so viele Werte stehen, wie unterhalb. Der Median ist aussagekräftig bei asymmetrischen Verteilungen der Werte und kleinen Stichproben.

Bsp.: In der Ex eines W-Seminars mit 13 Teilnehmern werden folgende Punktzahlen vergeben: 3, 4, 4, 5, 5, 6, 7, 8, 8, 10, 12, 13, 13. Der Median (also der Wert, der die Reihe so teilt, dass über und unter ihm die gleiche Anzahl an Werten stehen) wäre hier: 7 (darunter gibt es 6 Pkt.-Zahlen, darüber ebenfalls). Zum Vergleich: das arithmetische Mittel wäre 7,54.

Die Normalverteilung

Die meisten Werte sind in der Natur normalverteilt, d.h.: Betrachtet man zum Beispiel die Körpergröße von Kaninchen, so besitzen die meisten Tiere eine mittlere Größe, einige wenige sind sehr klein und einige wenige sind sehr groß:

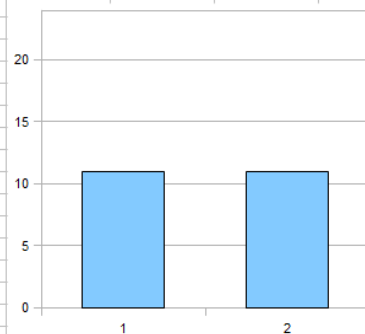


Die Standardabweichung $\pm s$:

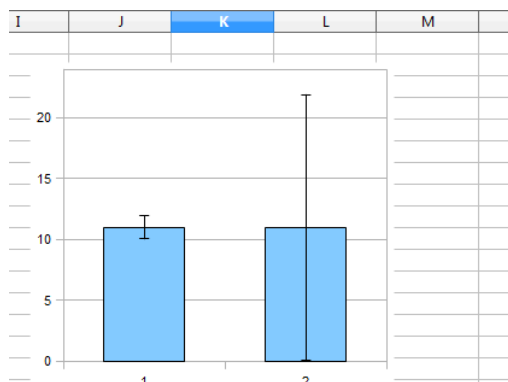
Mittelwerte alleine sind oft nicht besonders aussagekräftig.

Bsp.: Man untersucht die Länge von z.B. Schwanzfedern bei 2 Arten von Vögeln (Gruppe 1 & Gruppe 2 i. d. Abb.). In beiden Fällen ist der Mittelwert 11,0cm. Wenn man sich die Daten jedoch genauer anschaut, stellt man fest, dass die Schwanzfedern bei Vögeln der Gruppe 1 tatsächlich sehr oft 11,0 cm lang sind oder zumindest nahe an diesem Wert liegen. Bei Gruppe 2 dagegen so gut wie nie.

	A	B	C	D	E	F	G	H
1		z.B. Länge in cm						
2		Gruppe 1	Gruppe 2					
3		11	1					
4		10	24					
5		11	2					
6		12	0					
7		13	23					
8		10	2					
9		10	22					
10		11	3					
11		11	1					
12		11	23					
13		10	24					
14		12	7					
15		Mittelwert	11	11				
16		Standardabweichung	0.95	10.91				
17								



Um dies in einer Grafik deutlich zu machen, gibt man die Standardabweichung an. Sie ist ein Maß dafür, wie stark die Werte um den Mittelwert streuen. Je größer die Standardabweichung, umso weniger aussagekräftig ist der Mittelwert. Man kann die Standardabweichung mit so genannten „whiskers“ kennzeichnen:



Beim Median wird nicht die Standardabweichung, sondern das 1. und 3. Quartil angegeben. Das sind die Bereiche, in denen 50% der ermittelten Werte liegen.